Annual Review of Genetics

Power in Numbers: Single-Cell RNA-Seq Strategies to Dissect Complex Tissues

Kenneth D. Birnbaum

Center for Genomics and Systems Biology, New York University, New York, NY 10003, USA; email: ken.birnbaum@nyu.edu

Annu. Rev. Genet. 2018. 52:203–21

First published as a Review in Advance on September 7, 2018

The *Annual Review of Genetics* is online at genet.annualreviews.org

https://doi.org/10.1146/annurev-genet-120417-031247

Keywords

single-cell RNA-seq, trajectory analysis, rare cell types, pooled CRISPR, single-cell power analysis

Abstract

The growing scale and declining cost of single-cell RNA-sequencing (RNA-seq) now permit a repetition of cell sampling that increases the power to detect rare cell states, reconstruct developmental trajectories, and measure phenotype in new terms such as cellular variance. The characterization of anatomy and developmental dynamics has not had an equivalent breakthrough since groundbreaking advances in live fluorescent microscopy. The new resolution obtained by single-cell RNA-seq is a boon to genetics because the novel description of phenotype offers the opportunity to refine gene function and dissect pleiotropy. In addition, the recent pairing of high-throughput genetic perturbation with single-cell RNA-seq has made practical a scale of genetic screening not previously possible.

INTRODUCTION

A census—the collection of information at the level of the individual—can seem like a mundane exercise in data collection. But individual counts can reveal a qualitatively different picture than summary statistics. Consider one example from the US census. California has the tenth highest average household income, but, counting individuals, it is the state with the highest number of people in poverty.

Single-cell RNA-sequencing (RNA-seq) is a census of living tissue in which an arguably meaningful functional unit, the cell, is measured for a set of complex characteristics, the transcriptome. It permits a take-it-apart-and-see-what-it's-made-of approach that relaxes assumptions and constraints about tissue composition and organization. Cells may be removed from their tissue context, but they can be used to reconstruct the complex cellular environment from which they came.

Previously, when experiments sampled tissues for transcriptome analysis, often they were limited to collecting specific cell types that confounded many developmental stages or, alternatively, that represented specific developmental stages from a mix of cell types. With the ability to break tissues into their constituent cells, single-cell RNA-seq can be used to reorder cells into a step-by-step progression that mirrors their maturation in living tissue. With this resolution comes the ability to discover cell states not previously known or described. Indeed, using such techniques, single-cell RNA-seq studies have led to hypotheses of new precursors to specialized cells in blood, gut, and the nervous system (32, 78, 80).

Single-cell RNA-seq also permits new ways to measure tissues that were not practical with summary statistics from even highly targeted populations. For example, two recent single-cell RNA-seq studies showed that variance in cellular expression—a property that can be quantified only with discrete measurement of individuals—increases with age, suggesting an intriguing hypothesis for why organ function diminishes in older individuals (20, 47). In addition, cancer researchers point to the potential of single-cell RNA-seq to address questions about the heterogeneity of tumor environments and enable early diagnostic screens for rare precancerous cells in bodily fluids (6). The processing power of single-cell RNA-seq has recently been combined with CRISPR-mediated perturbations to reach a scale of genetic analysis that gives new power to test combinatorial interactions in genetic networks (1, 16, 17, 34, 49).

It is true that measurements are noisy in any given cell (8, 10, 25, 38). However, the power of the single-cell RNA-seq technique lies in discovering groups of discrete cells that share a complex transcriptomic phenotype. Although gene expression measurement in any given cell comes with high noise, the complexity of measurements provides the power to group cells in much the same way that a jigsaw puzzle piece in a detailed part of the image is easier to place because the complexity provides reinforcing information about where it belongs. The groups or clusters of cells then collectively identify a cell state, which was often unknown a priori, in the analysis phase. The more cells that define a state, the more power one has to identify the cellular state and the genetic mechanisms that distinguish that state.

The measurement of more cells is precisely what a clever set of breakthroughs has made possible through a reduction in cost and a dramatic increase in the scale of experiments. For example, the Tang et al. (75, 76) protocol originally examined just six cells in 2009. By 2015, plate-based automated liquid handling protocols enabled hundreds of cells to be sequenced (56, 65). These were followed by droplet-based protocols that broke the four-digit and five-digit barriers (39, 46) and included some efforts that reached six-digit cell sampling (17). In 2017, combinatorial barcoding of messenger RNAs (mRNAs) within cells laid out a straightforward approach to sequence hundreds of thousands of cells (13, 64) with the potential to reach the one million–cell mark and beyond in a single experiment (13). A notable aspect of this progress is that the experimental setup has gotten simpler and cheaper, permitting greater access to these techniques.

The emphasis in this review is on strategies that use single-cell RNA-seq to describe phenotype in ways not previously possible or practical. The goal is to provide the experimentalist with a broad view of the opportunities now available to explore phenotype and gene function, while paying mind to trade-offs in experimental design. The field is growing so rapidly that the emphasis here is on examples rather than comprehensive coverage of each subfield of single-cell transcriptomics. However, I point out some reviews in subfields for deeper reading on select topics. To anticipate where this fast-moving field may be headed, I briefly cover some new developments and lay out a potential future wish list.

A PRIMER ON PROTOCOLS AND POWER ANALYSIS

Current protocols fall under three general categories: plate-based, droplet-based, and combinatorial indexing techniques (detailed in 59). This progression also represents the increase in cell-throughput capacity over recent years. Each technique has some trade-off between increasing cell numbers and increasing technical noise. The two critical sources of noise are (*a*) the so-called dropout rates in which mRNAs fail to be reverse transcribed in the first step of complementary DNA (cDNA) synthesis, commonly referred to as sensitivity, and (*b*) infidelity in the amplification of cDNA that has been converted from an mRNA species. Once an mRNA is converted into cDNA, it must be amplified. The fidelity of amplification of an mRNA species turned cDNA is commonly termed precision; for example, biased amplification of an mRNA in one cell over another would lead to lower precision. The take-home message is that the plummeting cost per cell permits increasing statistical power that, to some extent, can overcome these two sources of noise.

Plate-Based Methods

Plate-based methods are almost always combined with fluorescence activated cell sorting (FACS) to place cells of interest into 96- or 384-well plates. The switching mechanism at the 5′ end of the RNA transcript (SMART)-seq2 protocol uses a template switching strategy during first-strand synthesis that favors full-length cDNA construction (60). For practical reasons, there is a trade-off between favoring full-length cDNAs and including unique molecular identifiers (UMIs), which help correct amplification noise by identifying PCR-generated duplicates. Single-cell RNA barcoding (SCRB) uses a similar approach but incorporates cell specific barcodes and UMIs on a poly(T) primer (67).

Cell expression by linear amplification and sequencing (CEL-Seq) and CEL-Seq2 also barcode individual cells within wells during first-strand synthesis but then amplify transcripts using in vitro transcription (IVT) rather than PCR (30, 31). The IVT approach was automated for liquid handling using the massively parallel single-cell sequencing (MARS-seq) protocol, which permitted use of UMIs at the plate, transcript, and cell levels (33). Plate-based protocols, in particular SMART-seq2, have the greatest sensitivity in terms of the number of detectable mRNAs per cell, but they also have the lowest throughput (relative to droplet-based and combinatorial barcoding methods) in terms of cell numbers.

Droplet-Based Methods

Throughput capacity broke an important barrier when protocols were freed from well plates. Two protocols, inDrop and Drop-seq, used barcoded beads with poly(T) primers containing cell-specific barcodes dispensed into microfluidic droplets (39, 46). Mixing rates are adjusted so that droplets with two cells are rare; for example, Drop-seq developers reported droplets with more than

one cell occurred from <1% to around 11%, depending on cell concentration (46). One important difference between Drop-seq and inDrop is that the former relies on PCR amplification using UMIs to reduce amplification noise while the latter uses an IVT amplification strategy to reduce the amplification noise. These protocols can routinely sequence tens of thousands of cells based on 3′ sequences, but they appear to capture fewer mRNAs per cell (i.e., to have lower sensitivity) compared to the plate-based methods (85).

In these methods, cells could be sorted based on markers and then loaded into one of the droplet-based instruments. However, the cell throughput is so high for the droplet-based methods that most experiments profile dissociated cells without any preselection. In addition, sorting would initiate a stress to which cells would have time to respond while being processed for inDrop or Drop-seq—a time that is typically 30 minutes or more depending on the number of cells being processed, although commercial droplet-based instruments have greatly reduced processing time.

Another important consideration is that cell capture rates in the droplet-based methods are low, typically around 10%, so an experimental setup in which the cells of interest are very rare cells and cannot be concentrated in high numbers could make this approach impractical. Commercial single-cell RNA-seq droplet devices, such as the 10x Genomics Chromium, which employs droplet-based processing, report higher capture rates of up to 65% and the rapid processing time of 10 minutes or less for thousands of cells in many cases (84). Many labs are using this option because it is highly convenient; however, its per-cell cost is higher than noncommercial setups and its reagent packages and protocols are less amenable to troubleshooting and customization.

Combinatorial Barcoding

In a recent addition to single-cell RNA-seq techniques as of this writing, combinatorial indexing techniques take a simplified approach by substituting droplets for the cell itself. These methods fix and permeablize cells, allowing all cDNA and library preparation steps to be done in the cell, which becomes its own self-contained reaction vessel. The method does away with any need for single-cell isolation because intact cells can be uniquely labeled in batches by sequential barcodes or indices (**Figure 1**).

In the single-cell combinatorial indexing RNA sequencing (sci-RNA-seq) protocol, batches of cells were pooled into 96-well plates by FACS, and a first-strand poly(T) primer with a well-specific barcode was incorporated in first-strand synthesis (13). All cells were then mixed and randomly resorted into wells for the PCR step during which they received a second well-specific barcode (**Figure 1**). With a 96 × 96-well barcode strategy, in a test case with a few hundred cells or nuclei, the rate of accidental barcode overlap was less than 1%, although this rate can be expected to vary with different pooling and barcoding strategies. The group used a 96 × 960 strategy (splitting second round labeling into 10 plates of unique well barcodes) to process approximately 42,000 single-cell RNA-seq profiles in a single experiment. Another method termed Split-seq used a similar strategy, but barcodes were introduced by the addition of 5′-ligated adaptors in three repeated labeling and resorting steps (64). Both papers show the feasibility of triple labeling (13, 64). For example, a 384 × 384 × 384 combinatorial system would make $>10^6$ single-cell RNA-seq barcodes experimentally feasible (13).

Furthermore, indexing and other techniques are compatible with nuclear mRNA profiling, which, as shown by several papers, gives robust readouts (27, 28, 40). This enables the analysis of tissues that are not easily dissociated or which show differences in cellular dissociation and have traditionally led to biases in the composition of cell preparations. These are early days for the combinatorial indexing techniques, but the dual advantages of accessing virtually any cell with nuclear profiling and of ease of processing with sequential barcoding have the potential to make

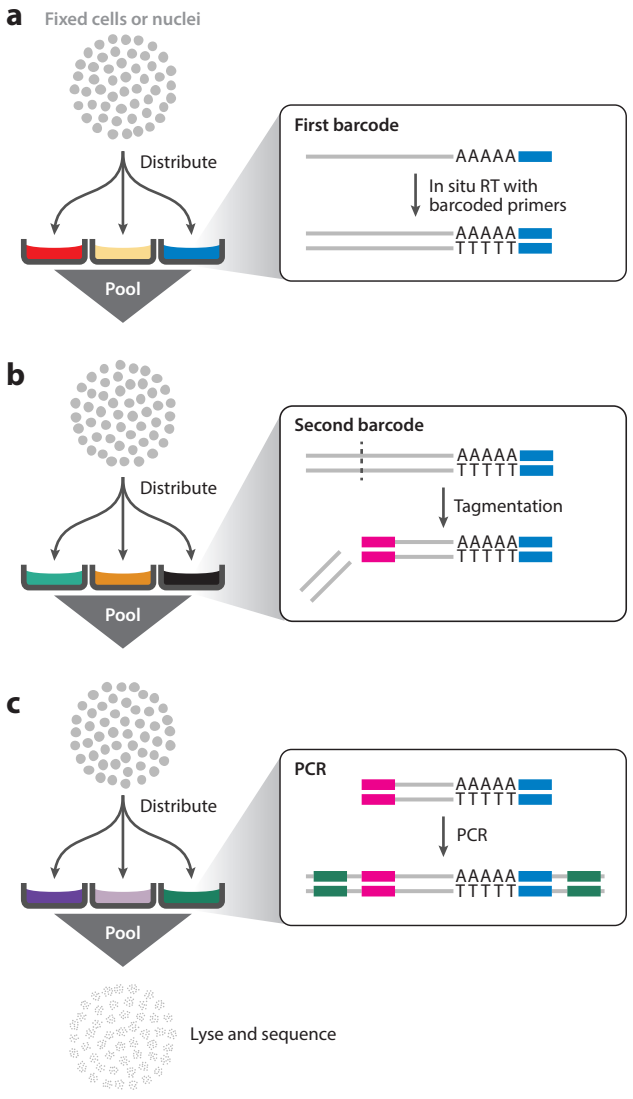

Figure 1

Strategy for combinatorial barcoding of cells in the sci-RNA-seq protocol. (*a*) Multiple fixed permeablized cells or nuclei are distributed in each well of a 96- or 384-well plate. The mRNA from each cell in a given well is reverse transcribed with a well-specific barcode (*blue*) using a poly(T) primer. (*b*) Cells are repooled and redistributed into wells once again for a transposase tagging and fragmentation. In the triple barcoding strategy, cells receive a second well-specific barcode with the transposase tag (*magenta*). The 3′ fragments with the poly(T) primers are specifically amplified in the next step. (*c*) Cells are repooled and redistributed again and then amplified with a primer that contains a third well-specific barcode (*green*). The example shows that cells landing in the right-most well received a sequential barcode of blue, magenta, green, where a 384 × 384 × 384 barcode strategy could yield millions of specific barcode combinations. Abbreviations: RT, reverse transcription; sci-RNA-seq, single-cell combinatorial indexing RNA sequencing. Figure credit: Ramin Rahni.

many new tissues accessible for single-cell RNA-seq. Still, the amount of mRNA in the nucleus is a fraction of the entire cell—already a challenging starting point—and fixation may further diminish mRNA yield per cell. More studies are needed to verify the early promising results based on single-nucleus profiling.

Cell Number and Power

Several recent benchmarking studies illustrated some of the trade-offs and opportunities that open up with increasing throughput. The benchmarking studies evaluated mRNA detection depth (sensitivity), read-count noise (precision), and sequencing costs per cell (74, 85). Ziegenhain et al. (85) simulated the likelihood of an experiment to detect a fixed set of relative differences in transcript abundance (which they termed power) given the observed dropout rates and noise associated with various plate- and droplet-based techniques. With a fixed number of cells to be profiled, benchmarking showed that each of the methods perform comparably (74, 85). For example, despite the added technical noise of the droplet-based technique as compared with SMART-seq, the former was shown to compensate for a higher noise profile by accounting for PCR amplification bias using UMIs (25, 85). Ultimately, in the comparison with fixed numbers of cells, plate-based methods such as SCRB that had moderate dropout rates and used UMIs to control for amplification noise performed best.

However, when cell numbers can be increased relatively easily, higher-throughput methods had the greater power per cost. For example, to detect 80% of true positives at a sequencing depth of 250,000 reads per cell, Drop-seq costs were the lowest at $690 for 254 cells (85). Still, other methods performed comparably; for gene detection purposes, practical matters such as having access to a core facility with a well-working protocol remain important considerations.

Two independent benchmarking studies concluded that one million reads per cell was a reasonable depth where detection power was still increasing at a good rate, at least for high sensitivity protocols (74, 85). However, they also showed that such depth is not always the most effective strategy. For example, the Ziegenhain et al. study (85) evaluated the number of cells needed to reach an 80% true-positive detection rate, given a range of simulated fold differences between two populations of cells. Using Drop-seq noise parameters, 135 cells were needed at 500,000 reads per cell, and only 99 cells were needed at one million reads per cell. However, neither of these strategies was as cost effective as the 254 cells at 250,000 reads per cell. Overall, the plummeting cost per cell often means that the most power per cost comes with sequencing more cells using high-throughput methods. Although the combinatorial indexing techniques were not tested in these benchmarking studies, one would expect them to have advantages that are similar to the droplet-based methods. One cautionary note is that benchmarking studies used sensitivities that are not always achieved in practice. For example, the noncommercial Drop-seq setup is low cost, but read complexity in a cell may typically result in the detection of only a few hundred transcripts; commercial versions of the protocol can be double the library preparation cost but have up to 10 times the sensitivity.

Thus, in terms of a choice of techniques: A limited ability to collect cells would favor high-sensitivity methods such as SMART-seq2, MARS-seq, or SCRB, whereas any simultaneous assessment for rare cell types and expression differences would favor high-throughput methods such as the high-sensitivity droplet-based or combinatorial indexing methods. In the former case, the ability to use UMIs should be a major consideration when full-length transcripts are not a critical experimental imperative. In the latter case, a given sequencing output trades off increased cell numbers to discover rare cell states and per-cell depth to detect transcriptional differences in cell groups. For example, if a tissue has relatively few cell types that are evenly represented,

benchmarking studies suggest that cell numbers can be sacrificed in favor of greater depth (perhaps 250,000 reads per cell or higher) to gain power in the detection of gene expression differences.

DISSECTING TISSUE PHENOTYPES: WHEN SINGLE CELLS MATTER

Heterogeneity and Rare Cell Types

Developmental studies often use specific markers or combinations of markers to isolate highly localized populations by FACS. However, the approach can become somewhat circular: Fluorescent markers have been chosen to reflect morphology, and markers are used to define the molecular state of cell types. One advantage of a deep census of cell identities is a comprehensive and unbiased inventory of cell types.

Such inventories were rapidly put to use to revisit cellular complexity in the brain, where single-cell analysis can address the complexity of the central nervous system (54), and in the immune system, where single-cell analysis can go beyond a finite set of cell surface markers typically used for classification (55). For example, about 2,400 dendritic cells of the human immune system were isolated using a set of known markers designed to capture the entire lineage (80). An unsupervised clustering approach identified a new potential precursor of conventional dendritic cells. Importantly, the new precursor comprised only about 0.02% of the targeted population of cells, so any signature from these cells would have been lost even in targeted pools. The unbiased cellular census enabled the identification of these cells by their unique signature among many hundreds of cells. It is one of many take-it-apart-and-see-what-it's-made-of approaches.

The scope of cell profiling has increased so dramatically that high-throughput single-cell RNA-seq exploratory approaches can now be performed on whole organisms (**Figure 2*d***). For example, the sci-RNA-seq combinatorial indexing technique was used to sequence all 762 cells of the *Caenorhabditis elegans* L2-stage larva by resampling each cell approximately 50 times for a total of 50,000 cells (13). Even at the relatively low per-cell read depth of approximately 20,000, the study could detect 8,613 genes. The vast majority of detected transcripts showed enrichment for specific cell types.

In another study, Drop-seq was used to generate a virtual reconstruction of the *Drosophila* (stage six) embryo, using about 6,000 cells (37). The authors sequenced to a similar per-cell depth and detected about 3,100 transcripts. The cell-by-cell virtual reconstruction permitted the mapping of transcriptional regulators, whose expression domain was not previously known, to just a few cells in specific regions of the embryo (37). Thus, even in this well-characterized model, the single-cell RNA-seq reconstruction of an organism could provide new descriptive detail, suggesting that such virtual single-cell maps could generate new resources in many model systems. One of the potentials of a single-cell RNA-seq whole-organ profiling approach is that no markers are needed to isolate cells so that single-cell RNA-seq may be used to rapidly create virtual expression maps in species that lack advanced tools such as an array of tissue-specific fluorescent reporters.

Although these studies opted for relatively shallow per-cell sequencing depths, they still detected a relatively high proportion of genes—especially in the high-cell number sampling scheme in *C. elegans* (13). Not all genes were detected and not all cells could be mapped to tissues using the limited set of known markers. However, these and other studies have demonstrated that the shallow-depth/high-cell number strategy is capable of mapping highly localized transcripts such as transcription factors. Benchmarking studies seem to suggest that higher per-cell depth may increase the power to identify localized transcripts. However, future studies aimed at developing community resources would seem to benefit from deeper per-cell sequencing to provide a comprehensive inventory of transcript abundance at each cell position.

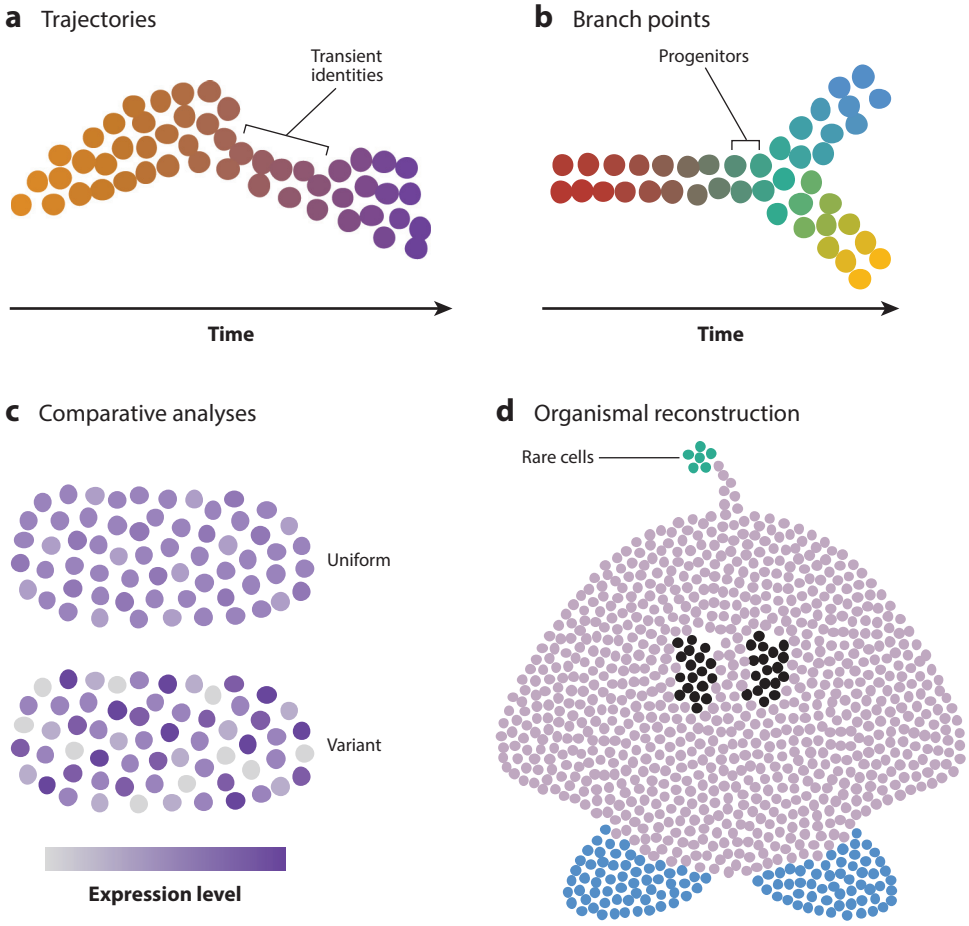

Figure 2

Applications of single-cell RNA-sequencing (RNA-seq) that refine tissue organization. (*a*) One prominent application of single-cell expression profiles is using complex expression patterns to order cells along a developmental trajectory. Sparse regions along the trajectory can signify highly transient cell states. (*b*) Branch point analysis reconstructs bifurcations in lineages and identifies cells that represent progenitor states. (*c*) Comparative studies at the single-cell level have the potential to dissect the effect of control versus treatment (or mutant) phenotypes and identify differences in the variance of gene expression between one or more groups of cells. (*d*) The throughput of single-cell RNA-seq studies has reached a scale that enables the profiling of entire tissues or organisms, permitting the discovery of rare cell types (*green*) or highly localized transcriptional domains not known a priori. Figure credit: Ramin Rahni.

Computational Strategies to Reconstruction

The noise inherent in single-cell RNA-seq readouts means that eyeballing a small number of markers to determine cell identity is highly error prone. Thus, a variety of computational approaches have arisen to classify cells and reconstruct their original positions. Unsupervised dimensional reduction techniques such as t-distributed stochastic neighbor embedding (t-SNE), which has been used to deal with sparse matrices such as outputs from single-cell RNA-seq (79), impose the fewest constraints on data. Other approaches use a priori information from known cell markers to classify cells by location or identity, employing various strategies to handle dropout and technical

noise. For example, the Seurat package accounts for the high dropout rate of single-cell RNA-seq by using the correlation of known markers to other genes in the data set (65). A putatively false-negative locational marker is then assigned to a cell based on whether its correlated markers are present with high enough frequency in the same cell.

Other approaches have used scoring to vote on cell identity based on the presence of known markers. False-negative markers from dropouts are presumed to be overwhelmed by the signal from other markers (2, 18, 37). This approach has a good deal of power when there is a large marker reference set. For example, the *Drosophila* embryo reconstruction used a confusion matrix, which scored positional matches based on the presence or absence of markers (37) (for a recent comprehensive review of spatial reconstruction techniques see 42).

Unsupervised clustering approaches do not make any assumptions about the a priori potential cell identities and, in theory, they are good approaches to discover cell states not previously identified. In practice, the dimensionality reduction techniques typically employed by unsupervised approaches are strongly influenced by dominant expression patterns so information that separates subtypes could be swamped. One approach to address this problem is to take subgroups of closely clustered cells and recluster the cells alone so that new expression trends distinguish the previously clumped subtypes (e.g., 13). Alternatively, supervised techniques such as confusion matrices rely on known markers and cannot detect new cell identities, although they have been used to detect chimeric identities (18, 19). These approaches are generally more sensitive and less susceptible to batch effects than unsupervised clustering, however, so the trade-off is between freedom to discover new cell states and sensitivity for detecting existing ones.

Mystery Cell States

In many cases, a particular cell state is known to exist, but reporters for the cell state may not be known or cannot specifically mark the cell. For example, the *AG2-G* gene was known to be involved in *Plasmodium falciparum* sexual reproduction and the switch away from persistent replication in the schizont phase (61). However, the early stage sexually-committed schizonts were difficult to isolate from their asexual counterparts, and, apart from *AG2-G*'s role, the gene expression program mediating this sexual commitment remained largely unknown. This gap was addressed with Drop-seq characterization of the malarial life cycle in wild-type *P. falciparum* and in a conditional *AG2-G* knockdown, in which the populations of sexual and asexual schizonts could be distinguished by identifying the cell cluster that was missing in a conditional *AG2-G* knockdown. This permitted identification of many candidate mechanisms that mediate critical epigenetic changes known to accompany the switch to a sexual life cycle (61).

In the study of regeneration, questions focus on the transitional state of cells during reprogramming when the expression patterns of cell identity markers are in the midst of repatterning. Single-cell RNA-seq has been used in both planaria regeneration and plant regeneration to characterize the transcriptomes of regenerating cells that were analyzed immediately after injury signals had begun to rapidly reshape cell identity (19, 83). In plants, regenerating root cells exhibited chimeric fates, as determined from reference profiles of adult cell types, but the chimeric identities most closely resembled a primed embryonic cell. In this study, single-cell profiles provided the first clue that regeneration recapitulated an embryonic sequence (19).

Cellular Trajectories

Many of the central questions of developmental biology are concerned with the process by which a naïve cell takes on a differentiated state. Which cells give rise to new identities? What signals

mediate cell fate decisions? How do cells make transitions in their regulatory states? Such questions are often difficult to address experimentally because available markers often convolute different cell stages or different cell types. However, if individual cells can be collected along a maturation gradient, single-cell RNA-seq offers a way to reconstruct the developmental continuum of specific cell types (**Figure 2*a***).

Some of the most frequently used computational tools to reconstruct trajectories include the Monocle (77), Wanderlust (7), and Waterfall (69) algorithms; (see 12 for a comprehensive review of trajectory techniques). Generally, these algorithms are designed for the pseudotemporal ordering of cells along a maturation gradient when true order has been lost in cellular dissociation. They cannot capture the true time scale of development because the same cell cannot be followed over time (hence, pseudotemporal ordering). These algorithms typically simplify the complex RNA readout with dimensionality reduction and then take a variety of approaches to order the cells de novo, usually without any a priori information. Many approaches to the problem work well, largely because developmental transitions are often dramatic and continual at the transcriptional level providing a lot of power.

One study used Waterfall to order neural cells from the quiescent stem cell state to the point at which they differentiated into daughter cells in the hippocampus. The program was also used to capture transitions in gene expression along the developmental continuum, revealing step-by-step transitions, for example, from metabolic processes such as glycolysis and glutathione and fatty acid metabolism in the quiescent state to activation of the cell cycle and the transcriptional and translational machinery (69). This study illustrates how single-cell profiles can help assess the step-by-step coordination of metabolic and cell identity networks.

With the ability to order cells along a developmental timeline, the opportunity arises to map the precise points where lineages bifurcate (**Figure 2*b***). Herring et al. (32) developed an algorithm called p-Creode, which uses graph theory and is designed to work with the densely-sampled cell data sets that are now possible with the emergence of high-throughput methods. Using 39,000 and 17,000 cells from the small intestine and colon, respectively, the study showed a branch point with two previously unknown origins of rare chemosensory tuft cells. Cells in the small intestine suggested a nonsecretory cell origin, whereas a different lineage was suggested for cells in the colon. These findings supported earlier mutational analyses which had identified differential function for these cells across the two organs in spite of the cell's presumed shared identity. Thus, initial uses of single-cell RNA-seq are beginning to reveal a better understanding of the early events that generate tissue complexity and identification of the cells from which new tissue types arise.

Indeed, knowing the earliest stages of a cell type's divergence from a progenitor could help to reconstruct the cellular environment and determine instructive signals in tissue formation. However, cells that are projected onto graphs of reduced dimensions often form an unresolved blob at the branch points that lead to lineage separation. Resolving the early progenitors of a specific branch within the blob would appear to be beyond the resolution of single-cell profiling, but La Manno et al. (41) use the signature of intronic reads from unprocessed mRNAs to predict the next cell state. The basis of the model is that the ratio of unprocessed to processed mRNAs is higher in early phases of a transcriptional induction and lower once transcription slows or shuts down. For example, the researchers model cells in a prebranch point undergoing neurogenesis in the hippocampus, and they predict that one set of cells in a blob is primed to enter one branch of the lineage and a second set of cells is primed to enter the other branch. The transcripts that contribute to the directional vectors can then be analyzed, and these two otherwise indistinguishable cell groups can be differentiated by possible inference of the distinct signals to which they are responding.

Inferring Instability from Sparsity

The snapshot of ordered cell states also provides some insight into the pace of development because density says something about the time cells spend in any given state. In one case, human adult spermatogonial stem cells were ordered using Monocle (77). Four distinct cell states were found with few intermediate cell states between the discrete groups (26), a finding that suggests rapid transitions during development into stable and discrete regulatory states. Another study of about 450 cells undergoing direct reprogramming from fibroblast to cardiomyocyte ordered this highly unsynchronized process using SLICER (44, 82). One low-density region corresponded to preinduced cardiomyocytes and led to the inference that these cells, which expressed both fibroblast markers and cardiomyocyte markers, represented an unstable state (44). The authors speculated that feedback mechanisms might make the transition state unstable, which would represent a potential barrier to reprogramming.

The ability to now profile entire organisms or tissues should also open up new opportunities to describe the relationship between independent lineages in the same tissue. Thus, approaches that can relax constraints about the history and developmental path of a lineage among cells can take greater advantage of the unbiased view of developing tissue that single-cell RNA-seq offers. Along those lines, single-cell topological data analysis (scTDA) develops a framework intended to relax assumptions about the path of developmental trajectories [e.g., trajectories may be circular or even discontinuous (63)]. In a proof of concept (63), cells from human preimplantation embryos were ordered both in a developmental gradient and into three distinct lineages—inner cell mass, early trophectoderm, and polar trophectoderm—without any reference data sets.

Direct Lineage Tracing

The techniques described previously order cells based on the similarity of their expression so there is no direct evidence of their clonal relationships. Gene editing techniques, in contrast, can be used to assess clonality by altering cells during development and reading out those alterations in single-cell RNA-seq profiles. Several protocols take this approach; for example, genome editing of synthetic target arrays for lineage tracing (GESTALT) and homing guide RNA (gRNA) use gRNA–targeted sequences to make continual edits in a DNA cassette that are read out by amplification of DNA or cDNA of the edited sequence (35, 48). In these strategies, a DNA cassette is typically introduced by stable transformation, creating a blank canvas for guide-directed editing to alter sequentially. The lineage history is inferred from the sequential editing of the DNA cassette, which is read out by sequencing and can be associated with a cell-specific barcode. In another strategy, Polylox employs a Cre recombinase–targetable cassette with many alternative recognition sites to generate millions of potential unique barcodes (57). When the readouts for the edited cassettes are transcribed, the technique is compatible with single-cell RNA-seq. For example, a method termed LINNAEUS (lineage tracing by nuclease-activated editing of ubiquitous sequences) targeted an endogenous red fluorescent protein reporter in zebrafish with CRISPR guides, whereupon both lineage information and transcriptome information could be read out together in the single-cell transcriptome profile (71).

Expression Variance Among Cells

Finally, there are some properties of tissue that were simply not feasible to measure without single-cell RNA-seq analysis, such as variance in gene expression (43, 53, 66, 81). While measures of variance need to exceed considerable technical noise, the sheer numbers of cells now being collected permit high confidence in the detection of differences in variance among cell populations (**Figure 2*c***).

Two separate papers, one on pancreas and another on immune cells, used single-cell profiles to independently describe similar trends in cellular variance with aging. In the pancreas, a total of 2,544 cells from differently aged donors were profiled, and it was found that islet endocrine cells from older individuals had a higher level of biological expression noise than cells from younger individuals (20). A second study on activated immune cells in mice found a similar trend of increased cellular variance with age (47). Interestingly, the activation of $CD4^+$ T cells in young mice and old mice did not show any average expression differences. However, while the activation of $CD4^+$ T cells led to the induction of a tightly controlled core program in young mice, older mice showed higher variance in activated cells. This suggests that expression variance is a key parameter to consider in the decline of organ function with age.

GENETIC SCREENS AND COMPARATIVE ANALYSES

If the applications described previously are powerful but largely descriptive tools, single-cell RNA-seq has also been put to use directly in genetic screens. Where pooled CRISPR screens have offered a way to perform large-scale knockout perturbations, single-cell RNA-seq has been a natural companion assay in that it offers a high-throughput way to assay complex phenotypes.

Pooled CRISPR Combined with Single-Cell RNA-Seq

In pooled CRISPR screens that were developed several years ago, thousands of different gRNAs are produced in high throughput and transfected at once into a cell suspension, typically a cell culture or primary cells. The readout is based on survival (positive screens) or failure to survive (negative or dropout screens) under some treatment such as a viral inoculation or an antibiotic (50). One feature lacking in these pooled perturbations was the specific effect of a given perturbation on the complex function of the cell—a gap that was filled with single-cell RNA-seq.

One problem with combined screens is that the genomic sequencing that would capture CRISPR-induced mutations and RNA-seq cannot be reliably assessed together in a given cell (at least for now). In addition, gRNAs are transcribed by RNA polymerase III and lack the poly(A) tail, which is used to generate most cDNA libraries for single-cell RNA-seq. To address these issues (**Figure 3**), Perturb-seq and CRISP-seq generated a gRNA construct that contained a fluorescent reporter and a selectable marker fused to a guide-specific barcode to enable selection for positive transformants and the readout of a guide-specific barcode using the cell-specific poly(T) primer (17, 34).

Other approaches using pooled CRISPR/single-cell RNA-seq include a version of Perturb-seq that uses CRISPR interference instead of direct genome editing (3). To streamline guide cloning, CRISPR droplet sequencing (CROP-seq) embedded the gRNA cassette in a repetitive sequence that is duplicated during lentiviral integration to generate the functional guide and readable tag in one step (16). In order to account for the fact that the same gRNA sequence can generate different DNA lesions, another technique, CRISPR-UMI, barcoded individual gRNAs, allowing for identification of clones stemming from a single gene editing event (49).

These early studies highlight the promise of single-cell RNA-seq to advance our understanding of the organization of genetic networks. For example, the Perturb-seq study targeted and grouped 24 transcription factors by their control of specific modules of downstream genes. In turn, modules could antagonistically regulate each other to influence the balance of different subtypes found among wild-type cells (17) in a manner reminiscent of so-called attractor states that mark specific cell identities (21). These early experiments also demonstrated how the scale of cell throughput could be used to explore combinations of knockouts to find additive, synergistic, and redundant

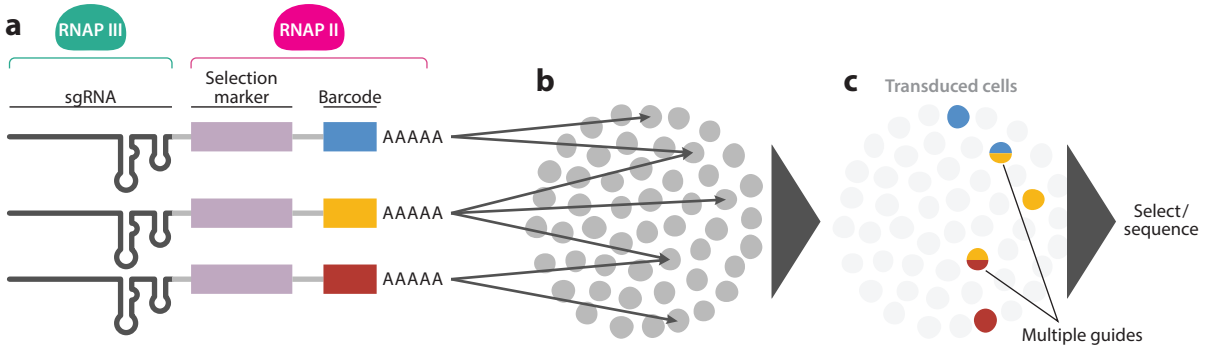

Figure 3

Scheme of pooled CRISPR/single-cell RNA-sequencing (RNA-seq) genetic screens. (*a*) A library of guide RNA constructs in a vector is prepared for transduction via viral vectors in animal cells or for transfection in other eukaryotic cells. Vectors typically contain an RNAP III–transcribed guide and an RNAP II–transcribed selectable marker of transformation with a barcode that identifies a specific guide (*blue*, *yellow*, or *red*). The guide barcode is read along with a cell-specific barcode via poly(T) priming to associate a guide with a cell. (*b*) Single or multiple guides may transduce or transfect a cell. (*c*) Multiple guides (*double-colored circles*) frequently generate higher order mutants. Abbreviations: RNAP, RNA polymerase; sgRNA, single-guide RNA. Figure credit: Ramin Rahni.

genetic interactions (17). Still, one open question for this technique is how future discoveries will reflect the structure and behavior of genetic networks in vivo.

Wild Type Versus Mutant: Batch Effects in Comparative Experiments

In genetic approaches, single-cell RNA-seq can also be used to dissect specific genotypes with a cell-by-cell inventory of alterations in transcriptional states. Such detailed views of phenotype could help address pleiotropy, for example, by finding the earliest stages of a developmental trajectory in which cellular transcriptomes begin to diverge from the wild type. However, the primary problem in such comparative experiments is the batch effects, which can be pronounced in single-cell experiments and which would confound any comparative analysis.

Experimental Pooling

The most straightforward approach to reducing batch effects in comparative experiments is to process all samples together from start to finish. If the genetic background of the two samples under comparison differs, polymorphisms can distinguish samples in an approach that was formalized by the developers of the demuxlet algorithm (36). The method can work with fairly minimal genetic divergence, and just a few reads from a cell can identify the cell's experimental origin, allowing the remaining reads to be assigned by their common cell barcode. For example, the demuxlet authors determined that >98% of reads could be accurately assigned to a sample with just 20 single-nucleotide polymorphism–containing reads having a minor allele frequency of 50% (36).

Another approach employs DNA-barcoded antibodies to ubiquitous and highly expressed cell-surface proteins, a technique termed cell hashing (73). Cells from two or more conditions are preincubated separately with a batch-specific barcoded antibody. The antibody barcode is then read out with the cell-specific barcode to assign a cell back into the appropriate sample. The authors showed highly concordant results with demuxlet when the two approaches were used together, but the cell hashing technique also offers a way to demultiplex batches when cells are too genetically similar to be treated by demuxlet.

Computational Correction

Computational correction is an alternative method of reducing batch effects in comparative experiments that do not fit the requirements for hashing or demuxlet. One correction method identifies similar cells in the two batches under comparison by their sharing of mutual nearest neighboring cells. Many pairs of cells are then used to generate a vector that realigns the two batches, but, importantly, the two batches do not need to have an identical composition of cell types (29). Another approach works directly on gene–gene correlations across the two data sets, finding covarying genes across the two data sets to align the two batches using canonical correlation analysis (11). The authors of this approach also showed that it could be used to align cells collected from different single-cell protocols or even different species (11).

RNA AND WHAT ELSE?

Multiple Readouts from Cells

What is next for this rapidly moving field? Many research labs are looking for ways to combine complex readouts at several levels of regulation. Single-cell readouts were first extended to open chromatin (9, 15) and then to bisulfite sequencing on single cells (22–24, 70). Increasing numbers of protocols read out properties from two or more levels of regulation, such as RNA-seq/genome (45), RNA-seq/bisulfite sequencing (4), and open chromatin/DNA-methylation using ATAC-seq and bisulfite sequencing/nucleosome phasing (62). Many of these protocols collect sparse data on any given cell, but, again, the key to these techniques is collecting many cells, grouping them in the analysis phase, and generating a collective signature for any given level of regulation.

Cell surface proteins can be detected in a sequencing readout with antibodies bound to DNA sequence tags (58, 68, 72). When quantifying different cell surface proteins, as opposed to the hashing technique described previously, multiple DNA–antibody barcodes can be used. The DNA barcodes are fused to antibodies so that they can be read out in single-cell sequencing libraries along with the cell-specific barcode. Barcode frequency then reflects the relative concentration of proteins within a cell. The approach was used on 13 well-characterized antibodies in immune cells and produced a quantitative readout of both mRNA and protein (72). Interestingly, readouts from cell surface proteins were much less prone to dropout effects because of their relatively high abundance. These combined techniques should provide new insights into how information is transferred, remembered, and altered from one level of regulation in the cell to another.

Prospectus

The study of anatomy at the organismal and cellular level laid the ground work for modern genetics by providing a basis to describe phenotypes. Just as live imaging by modern fluorescent microscopy provided a dramatic advance in understanding developmental dynamics, single-cell RNA-seq is revealing new ways to examine tissue organization and development—call it a neo-anatomical perspective. It may sound like an inglorious role for such a technological advance, but some of the most exciting applications of single-cell RNA-seq have revisited anatomy by taking tissues apart to reveal hidden cell states, developmental progressions, and the diversity of cellular states. In its descriptive role alone, single-cell RNA-seq has animated development by bridging the complex molecular and anatomical context of the cell.

Further advances that make stronger associations between the in vivo environment of the cell and complex single-cell readouts are needed. While lineage methods that have been developed to

retrace cellular position provide an important link between intact tissues and single-cell profiles, single-cell RNA-seq analysis would benefit these techniques by allowing an even tighter association of single-cell RNA-seq profiles and their specific tissue context. Some of the massively parallel in situ hybridization techniques such as multiplexed error-robust fluorescence in situ hybridization (MERFISH) (14, 51, 52) may address this issue, especially in parallel experiments with single-cell RNA-seq. In addition, single-cell RNA-seq methods that create local perturbations in tissue and track the response of cells in perturbed and adjacent cells at precise distances can provide more power to evaluate the influence of local signals on cell function. Ultimately, methods that combine the advances in live imaging (optogenetics, for example) with single-cell RNA-seq could fill an important gap. For example, one cell or a small group of cells could be light-activated for a perturbation after which perturbed cells and their neighbors could be subjected to single-cell RNA-seq. Fluorescent markers or postsequencing analysis could then identify the cell's position in relation to the perturbation.

The noise associated with single-cell RNA-seq has made many wary of its utility, but the sheer number of single-cell profiles now permitted has gone a long way toward addressing the noise problem. Benchmarking studies have demonstrated how cell numbers can compensate for much of the dropout and amplification noise. Still, despite the power brought by increasing cell numbers, some caution needs to be taken when planning experiments. The combination of high mRNA dropout rates and shallow sequencing can lead to a dearth of information about cells and the clusters they form. Low-information content could mean that an entire cluster of cells may show too few informative markers to be reliably identified, and the sequence output may not have enough complexity to resolve closely related cell types. The specific resources available for downstream analysis should be an important consideration in choosing a single-cell RNA-seq technique. For example, the number of available markers and the ability to detect them in the single-cell profiles are critical factors in making shallow sequencing useful, so a more limited proof-of-concept experiment may go a long way in evaluating the utility of a particular method. Another important consideration is whether the experimental focus is on classifying cells or identifying genes that show relevant expression patterns. The latter problem may often be better suited to a targeted collection of a smaller group of cells that use the high-sensitivity, deeper sequencing protocols. Any future improvements in lowering the dropout rates in the first steps of cDNA preparation will change this equation by increasing the information content of single-cell RNA-seq profiles at any sequencing depth (e.g., 5). Finally, the steady pace of decreasing cost that has permitted cell-throughput to increase may be reaching a limit because the cost of sequencing, which is not declining as fast as the cost of preparing cDNA libraries from cells, is an increasing proportion of per-cell cost.

SUMMARY POINTS

1. Single-cell RNA-sequencing (RNA-seq) profiles are revealing new details of developmental systems that were previously uncharacterized, allowing more precise assessment of gene function.
2. Power analyses show that the ability to profile many cells can address many issues that arise from technical noise in single-cell RNA-seq experiments.
3. The throughput of single-cell RNA-seq experiments is rapidly increasing and cost is declining. This is increasing power and raising the potential to quickly generate community resources such as cell-by-cell whole organism maps.

4. The combination of pooled CRISPR and single-cell RNA-seq offers a new approach for constructing genetic networks that includes large-scale combinatorial loss-of-function screening.
5. New experimental approaches are focused on profiling mRNA and other levels of regulation in single cells, including chromatin and protein. Single-cell RNA-seq with lineage analysis permits a greater capacity to connect cell profiles with in vivo development.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I acknowledge my plant anatomy professor, Ray Evert of the University of Wisconsin, who animated form and function of cells with his great passion for cell biology. I also acknowledge funding from the National Institutes of Health grant GM078279.

LITERATURE CITED

1. Aarts M, Georgilis A, Beniazza M, Beolchi P, Banito A, et al. 2017. Coupling shRNA screens with single-cell RNA-seq identifies a dual role for mTOR in reprogramming-induced senescence. *Genes Dev.* 31:2085–98
2. Achim K, Pettit JB, Saraiva LR, Gavriouchkina D, Larsson T, et al. 2015. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* 33:503–9
3. Adamson B, Norman TM, Jost M, Cho MY, Nunez JK, et al. 2016. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* 167:1867–82.e21
4. Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, et al. 2016. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* 13:229–32
5. Bagnoli JW, Ziegenhain C, Janjic A, Wange LE, Vieth B, et al. 2018. Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. *Nat. Commun.* 9:2937
6. Baslan T, Hicks J. 2017. Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nat. Rev. Cancer* 17:557–69
7. Bendall SC, Davis KL, Amir ED, Tadmor MD, Simonds EF, et al. 2014. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* 157:714–25
8. Brennecke P, Anders S, Kim JK, Kolodziejczyk AA, Zhang X, et al. 2013. Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* 10:1093–95
9. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, et al. 2015. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523:486–90
10. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, et al. 2015. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 33:155–60
11. Butler A, Satija R. 2017. Integrated analysis of single cell transcriptomic data across conditions, technologies, and species. bioRxiv 164889. **https://doi.org/10.1101/164889**
12. Cannoodt R, Saelens W, Saeys Y. 2016. Computational methods for trajectory inference from single-cell transcriptomics. *Eur. J. Immunol.* 46:2496–506
13. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, et al. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357:661–67

14. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. 2015. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348:aaa6090
15. Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, et al. 2015. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 348:910–14
16. Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, et al. 2017. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* 14:297–301
17. Dixit A, Parnas O, Li B, Chen J, Fulco CP, et al. 2016. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167:1853–66.e17
18. Efroni I, Ip PL, Nawy T, Mello A, Birnbaum KD. 2015. Quantification of cell identity from single-cell gene expression profiles. *Genome Biol.* 16:9
19. Efroni I, Mello A, Nawy T, Ip PL, Rahni R, et al. 2016. Root regeneration triggers an embryo-like sequence guided by hormonal interactions. *Cell* 165:1721–33
20. Enge M, Arda HE, Mignardi M, Beausang J, Bottino R, et al. 2017. Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell* 171:321–30.e14
21. Enver T, Pera M, Peterson C, Andrews PW. 2009. Stem cell states, fates, and the rules of attraction. *Cell Stem Cell* 4:387–97
22. Farlik M, Sheffield NC, Nuzzo A, Datlinger P, Schonegger A, et al. 2015. Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep.* 10:1386–97
23. Gravina S, Dong X, Yu B, Vijg J. 2016. Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol.* 17:150
24. Gravina S, Ganapathi S, Vijg J. 2015. Single-cell, locus-specific bisulfite sequencing (SLBS) for direct detection of epimutations in DNA methylation patterns. *Nucleic Acids Res.* 43:e93
25. Grun D, Kester L, van Oudenaarden A. 2014. Validation of noise models for single-cell transcriptomics. *Nat. Methods* 11:637–40
26. Guo J, Grow EJ, Yi C, Mlcochova H, Maher GJ, et al. 2017. Chromatin and single-cell RNA-seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. *Cell Stem Cell* 21:533–46.e6
27. Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, et al. 2017. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* 14:955–58
28. Habib N, Li Y, Heidenreich M, Swiech L, Avraham-Davidi I, et al. 2016. Div-seq: single-nucleus RNA-seq reveals dynamics of rare adult newborn neurons. *Science* 353:925–28
29. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. 2018. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* 36:421–27
30. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, et al. 2016. CEL-seq2: sensitive highly-multiplexed single-cell RNA-seq. *Genome Biol.* 17:77
31. Hashimshony T, Wagner F, Sher N, Yanai I. 2012. CEL-seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* 2:666–73
32. Herring CA, Banerjee A, McKinley ET, Simmons AJ, Ping J, et al. 2017. Unsupervised trajectory analysis of single-cell RNA-seq and imaging data reveals alternative tuft cell origins in the gut. *Cell Syst.* 6:37–51.e9
33. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, et al. 2014. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343:776–79
34. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, et al. 2016. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* 167:1883–96.e15
35. Kalhor R, Mali P, Church GM. 2017. Rapidly evolving homing CRISPR barcodes. *Nat. Methods* 14:195–200
36. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, et al. 2018. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* 36:89–94
37. Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, et al. 2017. The *Drosophila* embryo at single-cell transcriptome resolution. *Science* 358:194–99
38. Kim JK, Kolodziejczyk AA, Ilicic T, Teichmann SA, Marioni JC. 2015. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat. Commun.* 6:8687

39. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, et al. 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161:1187–201
40. Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, et al. 2016. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* 352:1586–90
41. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, et al. 2018. RNA velocity of single cells. *Nature* 560:494–98
42. Lee JH. 2017. Quantitative approaches for investigating the spatial context of gene expression. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 9:e1369
43. Li J, Klughammer J, Farlik M, Penz T, Spittler A, et al. 2016. Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. *EMBO Rep.* 17:178–87
44. Liu Z, Wang L, Welch JD, Ma H, Zhou Y, et al. 2017. Single-cell transcriptomics reconstructs fate conversion from fibroblast to cardiomyocyte. *Nature* 551:100–4
45. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, et al. 2015. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* 12:519–22
46. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, et al. 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–14
47. Martinez-Jimenez CP, Eling N, Chen HC, Vallejos CA, Kolodziejczyk AA, et al. 2017. Aging increases cell-to-cell transcriptional variability upon immune stimulation. *Science* 355:1433–36
48. McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J. 2016. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* 353:aaf7907
49. Michlits G, Hubmann M, Wu S-H, Vainorius G, Budusan E, et al. 2017. CRISPR-UMI: single-cell lineage tracing of pooled CRISPR–Cas9 screens. *Nat. Methods* 14:1191–97
50. Miles LA, Garippa RJ, Poirier JT. 2016. Design, execution, and analysis of pooled in vitro CRISPR/Cas9 screens. *FEBS J.* 283:3170–80
51. Moffitt JR, Hao J, Wang G, Chen KH, Babcock HP, Zhuang X. 2016. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *PNAS* 113:11046–51
52. Moffitt JR, Zhuang X. 2016. RNA imaging with multiplexed error-robust fluorescence in situ hybridization (MERFISH). *Methods Enzymol.* 572:1–49
53. Muraro MJ, Dharmadhikari G, Grun D, Groen N, Dielen T, et al. 2016. A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* 3:385–94.e3
54. Ofengeim D, Giagtzoglou N, Huh D, Zou C, Yuan J. 2017. Single-cell RNA sequencing: unraveling the brain one cell at a time. *Trends Mol. Med.* 23:563–76
55. Papalexi E, Satija R. 2018. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* 18:35–45
56. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, et al. 2015. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* 163:1663–77
57. Pei W, Feyerabend TB, Rossler J, Wang X, Postrach D, et al. 2017. Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* 548:456–60
58. Peterson VM, Zhang KX, Kumar N, Wong J, Li L, et al. 2017. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* 35:936–39
59. Picelli S. 2017. Single-cell RNA-sequencing: The future of genome biology is now. *RNA Biol.* 14:637–50
60. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R. 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10:1096–98
61. Poran A, Notzel C, Aly O, Mencia-Trinchant N, Harris CT, et al. 2017. Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature* 551:95–99
62. Pott S. 2017. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *eLife* 6:e23203
63. Rizvi AH, Camara PG, Kandror EK, Roberts TJ, Schieren I, et al. 2017. Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nat. Biotechnol.* 35:551–60
64. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, et al. 2018. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* 360:176–82
65. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. 2015. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33:495–502

66. Segerstolpe Å, Palasantza A, Eliasson P, Andersson E-M, Andréasson A-C, et al. 2016. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab.* 24:593–607
67. Semrau S, Goldmann JE, Soumillon M, Mikkelsen TS, Jaenisch R, van Oudenaarden A. 2017. Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. *Nat. Commun.* 8:1096
68. Shahi P, Kim SC, Haliburton JR, Gartner ZJ, Abate AR. 2017. Abseq: ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci. Rep.* 7:44447
69. Shin J, Berg DA, Zhu Y, Shin JY, Song J, et al. 2015. Single-cell RNA-seq with Waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* 17:360–72
70. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, et al. 2014. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 11:817–20
71. Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, et al. 2018. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* 36:469–73
72. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, et al. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* 14:865–68
73. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung B, et al. 2017. Cell "hashing" with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. bioRxiv 237693. **https://doi.org/10.1101/237693**
74. Svensson V, Natarajan KN, Ly L-H, Miragaia RJ, Labalette C, et al. 2017. Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* 14:381–87
75. Tang F, Barbacioru C, Nordman E, Li B, Xu N, et al. 2010. RNA-seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.* 5:516–35
76. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, et al. 2009. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6:377–82
77. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, et al. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32:381–86
78. Usoskin D, Furlan A, Islam S, Abdo H, Lonnerberg P, et al. 2015. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* 18:145–53
79. van der Maaten L, Hinton G. 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9:2579–605
80. Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, et al. 2017. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* 356:eaah4573
81. Wang YJ, Schug J, Won KJ, Liu C, Naji A, et al. 2016. Single-cell transcriptomics of the human endocrine pancreas. *Diabetes* 65:3028–38
82. Welch JD, Hartemink AJ, Prins JF. 2016. SLICER: inferring branched, nonlinear cellular trajectories from single cell RNA-seq data. *Genome Biol.* 17:106
83. Wurtzel O, Cote LE, Poirier A, Satija R, Regev A, Reddien PW. 2015. A generic and cell-type-specific wound response precedes regeneration in planarians. *Dev. Cell* 35:632–45
84. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8:14049
85. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, et al. 2017. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65:631–43.e4

Annual Review of Genetics

Volume 52, 2018

Contents

Errata

An online log of corrections to *Annual Review of Genetics* articles may be found at http://www.annualreviews.org/errata/genet